

A comparison between single and combined climate predictors successes on predicting South Australian spring rainfall

H. M. Rasel, M. A. Imteaz and F. Mekanik

Department of Civil and Construction Engineering, Faculty of Science, Engineering and Technology, Swinburne University of Technology, Melbourne, VIC 3122, Australia
E-mail: hrrasel@swin.edu.au

Abstract: The occurrence of rainfall over Australia is closely related with several key climate predictors, which are El Nino Southern Oscillation (ENSO), Indian Ocean Dipole (IOD) and Southern Annular Mode (SAM). Some researchers tried to explore the effects of these climate predictors in rainfall variability of different parts of Australia, particularly Western Australia, Queensland and Victoria. Nonetheless, clear association between separate or combined large-scale climate predictors and South Australian spring rainfall is yet to be established. Past studies showed that maximum rainfall predictability was only 20% with isolated/individual effects of ENSO and SAM predictors in this region. The present study further explored these hypotheses. For achieving better predictability of spring rainfall, this paper examined additional two important aspects; relationship between lagged individual climate predictors with spring rainfall as well as linked (multiple combinations of ENSO and SAM predictors) influence of these significant lagged climate indicators on spring rainfall prediction. Multiple regression (MR) modeling was used in this study. Two stations; Tarcoola and Mount Eba were chosen as case study of this region.

MR models with combine-lagged climate predictors (SOI-SAM based models) showed better generalization ability for both the model calibration (1957-2009) and model validation (2010-2013) periods for all the stations. Results also demonstrated that rainfall predictability significantly increased using combined climate predictors compared to predictability with individual effects of each predictor. The attained combined climate model predictabilities were 44% for Tarcoola and 40% for Mount Eba during calibration period. The predictabilities were significantly enhanced during model validation; the results are 94% for Tarcoola and 83% for Mount Eba. Whereas, the maximum rainfall predictabilities were limited to 33% and 30% respectively considering the effects of single climate predictors. Therefore, statistical analyses outlined the capabilities of SOI-SAM based combined climate predictors compared to their single/individual influences for predicting South Australian spring rainfall.

Keywords: *ENSO, SAM, multiple regression, correlation, rainfall prediction*

1. INTRODUCTION

Rainfall is an outcome of a continuous chain of phenomena involving different ocean currents, air temperature and atmospheric pressure. The ability to forecast rainfall several months or seasons in advance has been a goal of water resource managers for many decades. A reliable forecast of rainfall several months or seasons ahead can be beneficial for the management of land and water resources systems (Anwar *et al.*, 2008), particularly in Australia where the hydroclimatic variability is very high (Peel *et al.*, 2001). Many researchers have tried to establish the relationships between large-scale climate predictors with rainfall in different parts around the world (Niu 2012; Grimm 2011).

The variability of Australian rainfall has been linked to several dominant large-scale climate indices based on sea surface temperature (SST) and pressure differences anomalies originate from the Pacific, Indian and Southern Oceans. The climate predictors comprising the key climate indices; ENSO, IOD and SAM were used by many researchers in various past studies. Kirono *et al.* (2010); Risbey *et al.* (2009); Meneghini *et al.* (2007); Cai *et al.* (2001) investigated rainfall variability with selected predictors variable covering whole of Australia, while others are more focused on particular region such as South West Western Australia (Ummenhofer *et al.*, 2008 and England *et al.*, 2006), South Australia (Nicholls, 2010 and Evans, 2009) and East Australia (Verdon *et al.*, 2004). South Australia is one of the regions where clear association between separate or combined large-scale climate predictors and South Australian spring rainfall is yet to be established. Risbey *et al.* (2009) showed that maximum rainfall predictability was only 20% with isolated/individual effects of ENSO and SAM predictors in this region.

The more recent conducted researches especially on South Australian rainfall predictions including Chowdhury and Beecham (2013) and Cai *et al.* (2011), that analyzed the influence of simultaneous and separate role of single climate predictors at a time. Schepen *et al.* (2012) discovered that a strong relationship between simultaneous/concurrent climate predictors and rainfall does not essentially proved that there is also exists a lagged relationship, which is very important for rainfall predictions. In many cases the relationships of climate predictors and rainfall are more complex where only single predictor alone are unable to predict rainfall accurately. This combined relationship considering lagged-time effects of multiple climate predictors has not previously been attempted in South Australia. Keim and Verdon-Kidd (2009) claimed that south Australian rainfall variability is not determined by a single climate predictor itself, it is clear that complete rainfall variability could not explain by a single climate driver on this region.

Due to the geographical locations of South Australia, it is obvious that the SAM may also have much influence on rainfall variability in this region. Also, previous research has not considered lagged-time effects as well as multiple combinations (ENSO-SAM sets) of these key climate indicators at a time for assessing rainfall predictability. These two important reasons might be the main facts for showing very low predictability of South Australian rainfall. Thus, the objective of this study is to investigate the relationship of separate and combined lagged-time effects of ENSO and SAM predictors on South Australian spring rainfall, as a case study. Therefore, this study would be the extended and revised works conducted by Chowdhury and Beecham (2013), Cai *et al.* (2011) and Risbey *et al.* (2009). Moreover, it also distinguished from previous studies by predicting spring rainfall four consecutive years in advance using maximum possible lagged-time relationship of separate and combined climate indices. The outputs of the model were intended to be deterministic forecast means that can explore with quantitative measures which opposed to probabilistic forecast which can give results only above or below median value.

2. DATA AND STUDY AREA

The historical monthly rainfall data in millimeters from January 1957 to December 2013 were obtained from the Australian Bureau of Meteorology website (www.bom.gov.au/climate/data/). The climate predictor's data were obtained from Climate Explorer website (<http://climexp.knmi.nl/>). Two rainfall stations in South Australia; Tarcoola (TC) and Mount Eba (ME) were chosen as case study shown in the fig. 1.

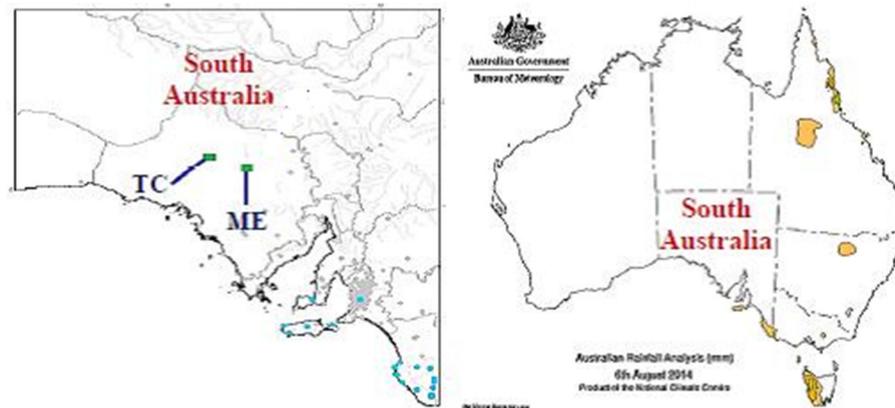


Figure 1. Map of the study area with selected locations (Source: www.bom.gov.au).

3. METHOD

To accomplish the aim of the study multiple regression (MR) modelling is used. MR analysis is a linear statistical modeling technique that allows finding out the best relationship between a variable (dependent, predicant) and several other variables (independent, predictor) through the least square method. The MR model can be expressed by a general equation 1 as following (Montgomery et al. 2001):

$$Y = b_0 + b_1X_1 + b_2X_2 + c \quad (1)$$

Where, Y is dependent variable (rainfall), X_1 and X_2 are 1st and 2nd independent variables respectively (lagged ENSO and SAM predictors), b_1 and b_2 are model coefficients of 1st and 2nd independent variable respectively, b_0 is constant, and c is the error. Multicollinearity verification among the predictors is the important stage of MR modeling. It occurs when the predictors are highly correlated that result in a dramatic change in parameter estimates in response to small changes in the data or the model. The Variance Inflation Factor (VIF) is used to ascertain the multicollinearity among the predictors. Lin (2008) found that a multicollinearity problem exists among the predictors if the VIF values greater than 5–10.

The performances of MR models were assessed by adopting several error indices and statistical performance test parameters which are widely used for the evaluation of prediction model. These includes: (i) Root mean square error (RMSE), (ii) Pearson multiple correlations (R), (iii) Willmott index of agreement (d), (iv) Variance Inflation Factor (VIF), (v) F-test and t-test. To overcome the shortcomings accompanying with R, MAE, and RMSE, Willmott (1981, 1982) developed the index of agreement (d), which were used for validating the developed forecasting models. Moreover, Willmott's measure is more appropriate for the investigation of model validation, where observed and model-predicted values need to be compared. The index of agreement (d) is expressed by the following equation 2:

$$d = 1 - \left(\frac{\sum(P_i - O_i)^2}{[\sum(|P_i - \bar{O}| + |O_i - \bar{P}|)^2]} \right) \quad (2)$$

Where, P_i is predicted or modeled value of the i th observation and O_i is the observed value of the i th observation. The optimum value of d is 1 called the better the model meaning that all the modeled values fit the observations (Willmott 1981, 1982). The SPSS statistical software was used to accomplish the single and multiple regression correlation tests. The correlations which are statistically significant at 1 and 5 % levels were considered in this study. The data were divided into two sets, years from 1957-2009 were used for calibration of the models. Later four years from 2010-2013 were selected as the out-of-sample test set to evaluate the generalization ability of the developed forecasting models.

4. RESULTS AND DISCUSSION

For evaluating the rainfall predictability, single/individual correlations between south Australian spring rainfall (S-O-N) at any year 'n' with lagged monthly values of ENSO and SAM climate predictors (NINO3, NINO4, NINO3.4 and SOI were chosen as ENSO predictors) from Dec_{n-1} - Aug_n ('n' being the year for which spring rainfall is predicted) were investigated. The correlations of rainfall with single predictor within the limits of statistical significance level and multicollinearity among the predictors were chosen for further MR analysis. It was observed that the maximum three months' (i.e. June, July and August) lagged SOI, Nino34 and Nino4 climate predictors have significant correlation with spring rainfall, whereas maximum five months of significant

lagged relation was found with SAM predictor. Moreover there is no further significant relationship for lags more than five months. Correlations of different lagged-time effects of individual climate predictors and SA spring rainfall are presented in table 1.

Table 1. Correlations of the different lagged time effects of individual climate predictors with spring rainfall

Station	Lagged climate indices								
	SAM _{Apr}	SOI _{Jul}	SOI _{Aug}	Nino34 _{Jun}	Nino34 _{Jul}	Nino34 _{Aug}	Nino4 _{Jun}	Nino4 _{Jul}	Nino4 _{Aug}
Tarcoola	0.28*	---	0.33*	-0.30*	-0.31*	---	-0.32*	-0.37**	-0.31*
Mount Eba	0.29*	0.30*	---	-0.33*	-0.36**	-0.32*	-0.36**	-0.42**	-0.39**

** : correlations are statistically significant at 1% level, * : correlations are statistically significant at 5% level

The results showed good consistency with the previous findings of Chowdhury and Beecham (2013); Cai et al. (2011); Nicholls (2010) and Menegnini et al. (2007). It is seen that Tarcoola is showing better correlations of SOI than Mount Eba station. In Mount Eba, both Nino34 and Nino4 predictors are showing better correlations with spring rainfall compared to Tarcoola. Spring rainfall is significantly influenced by SOI, particularly in July and August, but no influence in June’s rainfall like other phases of ENSO indicators. Moreover, the spring rainfall is also found significantly correlated by SAM driver in April in this region. The other phases of ENSO that is Nino34 and Nino4 having much influence compared to other climate predictors.

The combined lagged-predictor model sets were organized based on the significant single and separate lagged relationship obtained in table 1 with ENSO and SAM climate predictors. Then, ENSO-SAM based combined predictor model input sets were used for further multiple regressions modelling in order to investigate the predictability of spring rainfall with combined climate model. F-test and t-test were conducted to evaluate the significance levels of the MR models and the regression coefficients. Among the developed models the ones that follow all the limits of statistical significance level were selected, the models having lower error were chosen as the best model for rainfall prediction.

The results in table 2 shows the summary of the best developed MR models for the two different stations mentioning the values of regression coefficients and Variance Inflation Factor (VIF). The VIF indicators for the selected models are one and thus there is no multicollinearity problems exist among the predictors. Moreover, the D-W test statistics fall around a value of two which elucidate that the residuals of the predicted models has no autocorrelation and they are independent that confirmed the goodness-of-fit of the models.

Table 2. Summary of the best MR models (significant correlations at least 5% level)

Station	Models	Coefficient					VIF
		Const.	SOI _{Jun}	SOI _{Jul}	SOI _{Aug}	SAM _{Apr}	
Tarcoola	SOI _{Aug} -SAM _{Apr}	16.19	---	---	4.64	1.95	1.00
Mount Eba	SOI _{Jul} - SAM _{Apr}	11.68	---	3.36	---	1.50	1.00

Various performance evaluation statistics such as RMSE, multiple regression correlations (R) and index of agreement (d) of the best MR models during the calibration periods for the two studied regions are shown in the following table 3. SOI-SAM based models demonstrated statistically significant with better predicting ability for south Australian spring rainfall, with R= 0.44 for Tarcoola and 0.40 for Mount Eba.

Table 3. Performances of the MR models during calibration periods

Station	Models	Results for calibration period (1957-2009)		
		R	RMSE	d
Tarcoola	SOI _{Aug} -SAM _{Apr}	0.44	10.06	0.55
Mount Eba	SOI _{Jul} -SAM _{Apr}	0.40	8.24	0.50

After calibrating the models an out-of-sample tests were carried out on the years from 2010-2013 to evaluate the future rainfall predicting ability of the developed models. Similarly the RMSE, multiple regression correlations (R) and index of agreement (d) statistics were calculated during the model validation periods for the same regions, the results are shown in the table 4. MR model in validation stage is showing better generalization ability for all stations; however the ability of MR models to forecast out-of-sample sets improves significantly for Tarcoola with R= 0.94. The RMSE values of the testing sets for MR models are also compatible compared to the calibration stage. Mount Eba is having higher index of agreement value (d) in validation sets than other region; however the R value is bit smaller than other stations. All the ‘d’ values in the validation sets are more than 0.50 confirming that the SOI-SAM based combined climate predictor models are capable of predicting south Australian spring rainfall.

Table 4. Performances of the MR models during validation periods (out-of-sample test sets)

Station	Models	Results for validation period (2010-2013)		
		R	RMSE	d
Tarcoola	SOI _{Aug} -SAM _{Apr}	0.94	11.99	0.51
Mount Eba	SOI _{Jul} -SAM _{Apr}	0.83	12.07	0.55

It can be seen from the table 4 that after combining the climate predictors, it significantly increased the rainfall predictability up to 94% for Tarcoola with SOI_{Aug}-SAM_{Apr} model. Considering the correlation, statistical performance parameters and lower error, the best model developed in Tarcoola among the predicted models is shown in following equation 3:

$$\text{Rainfall} = 4.64 \times \text{SOI}(\text{Aug}) + 1.95 \times \text{SAM}(\text{Apr}) + 16.19 \tag{3}$$

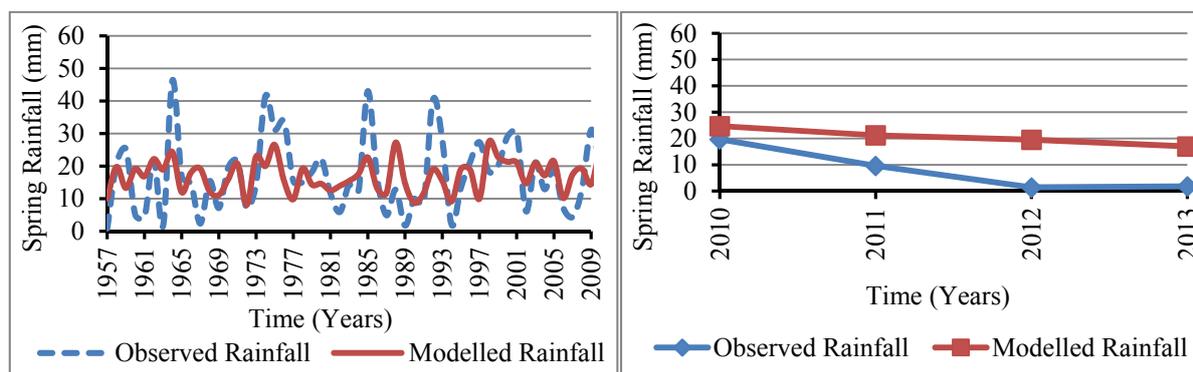


Figure 2. Performances of the best predicted model (1957-2009: calibration, 2010-2013: validation period)

Figure 2 shows the performances of the best models. Multiple regressions with combined-lagged climate predictors were able to model the observed rainfall in a way that the models follow the pattern of rainfalls several years in advance with very good accuracy.

The simulation results as well as various evaluations of statistical significances demonstrated that the developed SOI-SAM based combined climate models are capable of predicting spring rainfall. The developed MR models in validation period are showing an overestimation of the actual observed data series that means other climate influence may also be involved and should be taken into account for more accurate predictions in future studies.

5. CONCLUSION

Study focused on investigating the influence of single and combined lagged ENSO and SAM climate predictors on South Australian spring rainfall. NINO3, NINO4, NINO3.4 and SOI were chosen as ENSO climate predictors. The correlations of rainfall with single predictor within the limits of statistical significance level and multicollinearity among the predictors were chosen for further MR analysis. It was observed that the maximum three months (i.e. June, July and August) lagged SOI, Nino34 and Nino4 climate predictors have significant correlations with spring rainfall, whereas maximum five months significant lagged relation was found with SAM predictor. Results also demonstrated that the highest and significant correlations were achieved between

spring rainfall and single climate indices with maximum of three month lagged for ENSO and five months lagged for SAM predictors. Moreover there is no further significant relationship for lags more than five months. It is seen that Tarcoola is showing better correlations of SOI than Mount Eba station. In Mount Eba, both Nino34 and Nino4 predictors are showing better correlations with spring rainfall compared to Tarcoola. Spring rainfall is significantly influenced by SOI, particularly in July and August, but no influence in June's rainfall has been found like other phases of ENSO indicators. Moreover, the spring rainfall is also found significantly correlated by SAM driver in April in this region. The other phases of ENSO that is Nino34 and Nino4 having much influence compared to other climate predictors. Furthermore, the combinations of significant lagged predictor variables were used in MR modeling to investigate the predictability of spring rainfall. The RMSE values of the validation sets for MR models are compatible compared to the calibration stage. Mount Eba is having higher index of agreement value (*d*) in validation sets than other region; however the R value is bit smaller than other stations. All the '*d*' values in validation sets are larger than 0.50 confirming that SOI-SAM based combined climate predictor models are capable of predicting south Australian spring rainfall with good accuracy.

Multiple regression (MR) models with combine-lagged climate predictors (SOI-SAM) showed better generalization ability for both the model calibration and validation periods for all the stations. Results also demonstrated that rainfall predictability significantly increased using combined climate predictors compared to predictability with individual effects of predictors. The attained combined model predictabilities are 44% for Tarcoola and 40% for Mount Eba during calibration period. The predictabilities were significantly enhanced during model validation; that are 94% for Tarcoola and 83% for Mount Eba. In general, SOI-SAM based models showed good generalization ability for both the stations. Therefore, SOI-SAM based combined predictor models demonstrated statistically significant relationships with good forecasting ability for south Australian spring rainfall. The statistical analyses outlined the capabilities of combined climate predictors in compared with their single/individual influences for predicting spring rainfall using multiple regressions modeling. Further investigations of this method with other rainfall stations may be added to suggest a generalize model for rainfall forecasting which will be covered in future studies.

REFERENCES

- Anwar, M. R., Rodriguez, D., Liu, D. L., Power, S., O'leary G. J. 2008. Quality and potential utility of ENSO-based forecasts of spring rainfall and wheat yield in south-eastern Australia. *Australian Journal of Agricultural Research* 59: 112–126.
- Cai, W., Whetton P., Pittock, A. B. 2001. Fluctuations of the relationship between ENSO and northeast Australian rainfall. *Climate Dynamics* 17: 421–432.
- Cai, W., van Rensch, P., Cowan, T., Hendon, H. H. 2011. Teleconnection pathways of ENSO and the IOD and the mechanisms for impacts on Australian rainfall. *Journal of Climate*, 24(15), 3910-3923.
- Chowdhury, R. K., Beecham S. 2013. Influence of SOI, DMI and Niño3.4 on South Australian rainfall. *Stoch. Env. Res. Risk A.* 27(8): 1909–1920.
- England, M. H., Ummenhofer, C. C., & Santoso, A. 2006. Interannual rainfall extremes over southwest Western Australia linked to Indian Ocean climate variability. *Journal of Climate*, 19(10), 1948-1969.
- Evans, A. D., Bennett, J. M. & Ewenz, C. M. 2009. South Australian rainfall variability and climate extremes. *Climate Dynamics*, 33, 477-493.
- Grimm, A. M. 2011. Interannual climate variability in South America: impacts on seasonal precipitation, extreme events, and possible effects of climate change. *Stoch Environ Res Risk Assess* 25(4): 537–554.
- Kiem, A. S. and Verdon-Kidd, D. C. 2009. Climatic drivers of Victorian stream flow: Is ENSO the dominant influence. *Australian Journal of Water Resources*, 13(1), 17-29.
- Kirono, D. G. C., Chiew, F. H. S., Kent, D. M. 2010. Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrological Processes*, 24(10), 1237-1247.
- Lin, F. J. 2008. Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Quality & Quantity*, 42(3), pp. 417-426, 2008.
- Meneghini, B., Simmonds, I., Smith, I. N. 2007. Association between Australian rainfall and the Southern Annular Mode. *International Journal of Climatology*, 27(1), 109-121.
- Montgomery, D. C., Peck, E. A., Vining, G. G. 2001. "Introduction to linear regression analysis." Third edition, John Wiley & Sons, New York, USA.
- Nicholls, N. 2010. Local and remote causes of the southern Australian autumn-winter rainfall decline, 1958–2007. *Climate Dynamics*, 34(6), 835-845.
- Niu, J. 2012. Precipitation in the Pearl River basin, South China: scaling, regional patterns, and influence of large-scale climate anomalies. *Stoch Environ Res Risk Assess*. doi:10.1007/s00477-012-0661-2.
- Peel, M. C., McMahon, T. A., Finlayson, B. L., Watson, T. A. 2001. Identification and explanation of continental differences in the variability of annual runoff. *Journal of Hydrology* 250: 224–240.

- Risbey, J. S., Pook, M. J., McIntosh, P. C., Wheeler, M. C., Hendon, H. H. 2009. On the remote drivers of rainfall variability in Australia. *Monthly Weather Review*, 137(10), 3233-3253.
- Schepen, A., Wang, Q. J., Robertson, D. 2012. Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *Journal of Climate*, 25(4), 1230-1246.
- Ummenhofer, C. C., Sen Gupta, A., Pook, M. J., England, M. H. 2008. Anomalous rainfall over southwest Western Australia forced by Indian Ocean sea surface temperatures. *Journal of Climate*, 21, 5113-5134.
- Verdon, D. C., Wyatt, A. M., Kiem, A. S., Franks, S. W. 2004. Multidecadal variability of rainfall and streamflow: Eastern Australia. *Water Resources Research*, 40, W10201.
- Willmott, C. J. 1981. On the validation of models, *Physical Geography*, 2(2):184–194.
- Willmott, C. J. 1982. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, 63(11): 1309-1313.